

Defining Information: Citation Analysis and Co-Citation Analysis

One of my colleagues (a crony with a smattering of education) asked me what information was and why it mattered if we had a definition for it. I did not have a clue, and this was quite embarrassing. I decided to research this topic:

Introduction

Information is sort of like pornography; you cannot describe it but you know it when you see it.

Our reason for needing a definition is easy. We need at least a rough definition so we will know when we are talking about the same thing. It is sort of like a definition of “space,” which I clarified many years ago: “Space is a place you’ve got where you can put something, but you just haven’t put anything there yet.” Once you know what your definition is, you will have much less trouble talking about it.

My colleague by innuendo asserts that information is important, and I will stipulate to that. I will presume that information is what smart, educated people talk and write about. Since I am a reference librarian, I will narrow my hunt to mostly well-known Library Information Sciences (LIS) educators.

All extant LIS educators seem to agree that Patrick Wilson was a great educator. Surely great educators today are writing about at least some of the same basic concepts that Patrick Wilson wrote about and, when they do, surely they cite old Pat. If I can figure out what these educators are writing about, at the same time they are citing Patrick Wilson, I bet I will know what is important.

LIS educators

Patrick Wilson

Patrick Wilson (1968) mused: “There seem to me to be two quite distinct sorts of things that deserve to be called exercises of bibliographical control.” As I read his book, I mused: If bibliographic control is to yield “power over writings,” each form of bibliographic control (exploitative & descriptive) must match the searcher’s efforts.” (Wilson, 1968) Scholars have continued to write around this concept while citing Wilson.

Richard P. Smiraglia

In his study, Smiraglia (2007) sets the bar a mite higher for his colleagues by using citation analysis and co-citation analysis as a domain analysis tool for analyzing the corpus of literature that cites Patrick Wilson’s (1968) Two Kinds of Power: An Essay in Bibliographical Control. “Bibliographic Co-Citation is a popular similarity measure used to establish a subject similarity between two items.” (Class CoCitation, 2010) Smiraglia uses *Web of Science*, and his purpose is to ascertain what themes or concepts were dealt with by the authors involved. By using maps and diagrams generated by *Web of Science*, and eventually cutting-and-pasting parts of two maps,

Smiraglia culminates his work into his Figure 5 (Chart A below). (I'm not sure if Smiraglia used Elmer's or Schotch.) (Smiraglia ,2007)

You will notice (below) the solid lines that delineate the two clusters (1) Knowledge Organization (“KO”) and (2) Information Retrieval (“IR”); and a dashed line that surrounds the cluster Smiraglia calls (3) Library-and-Information Science. Library-and-information science is totally incorporated within the two other research streams.

Chart A
Copy of Smiraglia's (2007) Figure 5
 (Nice work, old chap.)

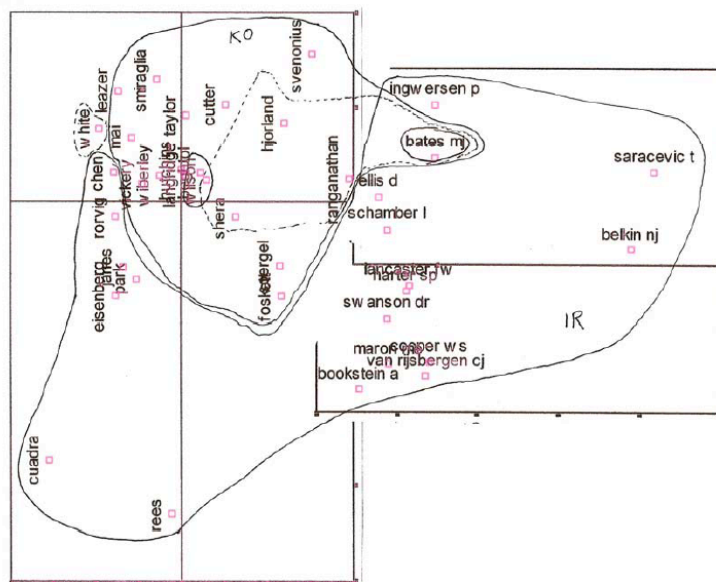


Figure 5. Clusters Demonstrated

(Smiraglia, 2007)

You will notice (above) that the three areas have a common overlap that implicates author Bates, who remains central to all research streams in this map. Smiraglia submits (I think he is right.) that this represents a core set of authors who provide a value system at the meta-level in information science that is shared by the two research streams that are clearly prevalent.

Smiraglia (2007) shows that the two concepts of Knowledge Organization and Information Retrieval are core values, or at least hot topics. Smiraglia (2007) points out that there are further degrees of specialization. “Knowledge organization here is principally the construct of subject access and relevance; information retrieval is principally the province of knowledge representation and relevance. Knowledge representation at the conceptual level is central to knowledge organization; knowledge representation at the application level is essential for information retrieval. Both are dependent on the concept of relevance.”

As I follow Smiraglia (2007), the conceptual level is where the cataloger works and the application level is where the searcher works—in the simplest terms possible.

Chart B
Chart to illustrate the above paragraph

	Conceptual level relation (Cataloger works here)	Application level relation (Searcher works here)
1.0 <u>Knowledge organization</u>	Knowledge representation is central to knowledge organization.	
1.1. Subject access		
1.2. <u>relevance</u>	Knowledge representation is dependent on the concept of relevance.	Knowledge representation is essential for information retrieval.
2.0 <u>Information retrieval</u>		
2.1. <u>knowledge representation</u>		
2.2. <u>relevance</u>		Knowledge representation is dependent on the concept of relevance.

Smiraglia cleverly points out that though the concept of *relevance* remains unarticulated, scholars operate within Wilson's (1968) exploitative domain by some unexplained process. (Smiraglia, 2007, page 11) Wilson (1968) would be proud, but most librarians are not getting it. For example, JISC Digital Media, cites the four categories of metadata from the Metadata Encoding and Transmission Standard or METS: (1) structural metadata, (2) descriptive metadata, (3) administrative metadata, and (4) technical metadata; but no exploitative metadata. (JISC, 2010) If Patrick Wilson were still around, he would probably be so upset that he would neglect to invite some scholars to the next pachanga.

Blog - Some things read this week, 27 April – 3 May 2008

From reviewing this blog, it is apparent that educators, or at least future librarians, are still reading Patrick Wilson's (1968) book and his 1983 book, *Second-Hand Knowledge. An Inquiry into Cognitive Authority*. It is difficult to tell who is writing what on this blog.

I may well have to write about this later. Seeing as it is bibliometric I need to comment on why I am more accepting of this piece than, say, Cronin and Meho above. There is much more to this piece though, for me, than its bibliometric issues. That is, it is far more meaningful for me as a whole.

Short, 13-pages with citations. *Well worth reading* as an example of domain analysis around "a classic work" [in our own field even].

The short answer as to why this sits better with me is because in one sense it validates much of my reading of the last 4+ years. The literature described by Smiraglia is a good description of what I have spent my time on for a while now. It is one [good] description of my view of the literature. It *validates* me.

It ain't exactly rational, but its true.
(Linder, 2008)

Cronin and Meho

The Linder (2008) blog discusses Smiraglia (2007) and Cronin and Meho's (2007) work. Cronin and Meho's (2007) work is not on point with Smiraglia (2007), but may be used for impeachment purposes; and while inferences can be had that could impeach Smiraglia (2007), the basis of the inferences could be supportive as well.

The authors describe a large-scale, longitudinal citation analysis of intellectual trading between information studies and cognate disciplines. The results of their investigation reveal the extent to which information studies draws on and, in turn, contributes to the ideational substrates of other academic domains. Their data show that the field has become a more successful exporter of ideas as well as less introverted than was previously the case. (Cronin & Meho, 2007)

Citation analysis is a powerful means of mapping the flow of ideas between specialty groups, disciplines, and nation states ... The matrices and maps produced by bibliometricians and others can be used to demonstrate the relative impact and perceived utility of research. (Cronin & Meho, 2007)

Summary

I now know what information is, as set out in Chart B. This appears to me to be the basic building blocks for all other information. Whether it is data, facts, ultimate facts, conclusions, or the existence of concepts, this basic information must exist to push the rest. This is part of what God created.

...

References

- JISC Digital Media. (2010, January 07) *Metadata and digital images*. Digital Media. Viewed February 22, 2010 at <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/metadata-and-digital-images/>
- Linder, Mark. (2008, May 3) *Some things read this week, 27 april – 3 may 2008*. Habitually probing generalist was off the mark. Viewed February 22, 2010 at <http://marklindner.info/blog/2008/05/06/some-things-read-this-week-27-april-3-may-2008/>
- Smiraglia, Richard P. (2007, May 10-12). *Two kinds of power: insight into the legacy of patrick wilson*. IN: CAIS 2007. Information Sharing in a Fragmented World: Crossing Boundaries. McGill University, Montreal, Quebec. Proceedings Editors: Clément Arsenault and Kimiz Dalkir . Poor link http://www.cais-acsi.ca/proceedings/2007/smiraglia_2007.pdf The citation was found at <http://www.db.dk/bh/core%20concepts%20in%20lis/articles%20a-z/wilson,%20Patrick.htm>
- Wilson, Patrick. (1968). *Two kinds of power: an essay in bibliographical control*. California library reprint series ed. Berkeley: University of California Press.